

Spectral marks for qualitative discriminant analysis

Olga M. Conde^{*}, Lucía Uriarte, Pilar B. García-Allende, Ana M. Cubillas, F. Anabitarte, Jose M. Lopez-Higuera
Photonics Eng.Group, Dep. TEISA, University of Cantabria, Avda. Castros s/n, Santander, Spain

ABSTRACT

In this paper, a method for the automatic qualitative discrimination of liquid samples based on their absorption spectrum in the ultraviolet, visible and near-infrared regions is presented. An alternative implementation of conventional spectrum matching methodologies is proposed working towards the improvement of the response time of the discrimination system. The method takes advantage of not making assumptions on the probability density function of the data and it is also capable of automatic outlier removal. Preliminary discrimination results have been evaluated on the classification of different oil samples from seeds and olives. The system here proposed could be easily and efficiently implemented in hardware platforms, improving in this way the system performance.

Keywords: qualitative discriminant analysis, absorbance spectrum, spectrum matching methods

1. INTRODUCTION

The purpose of discriminant analysis techniques is to classify samples into well-defined groups or categories based on a training set of similar samples [1]. A good discriminant algorithm is able to learn what the spectrum of a sample looks like by training it with spectra of the same material. There are a variety of algorithms to perform this spectral matching methodology readily offered in many commercially available programs. We can employ the Euclidean distance method, the vector correlation method, K-nearest neighbors [2], cluster analysis [3], PCA (Principal Component Analysis) factorial discriminant analysis [4], SIMCA [5], BEAST [6], etc. Always the trade-off time efficiency versus pattern recognition success figures has to be evaluated. Some of them are really time efficient but assume more error in the classification. In this work we propose a new alternative focused in the improvement of both indicators. This proposal will associate a figure with a specific spectral mark to each material category to be discriminated. The main advantage of the method is the really fast implementation of the testing stage of a typical pattern recognition structure.

2. SPECTRAL MARK CONCEPT

The initial concept of “spectral marks” is pretty close to the concept of direct spectral matching [7] but it differs in the generalization degree of the spectra comparison. Initially, it departs from the measurement of the spectrum of the transmitted light once it traverses a liquid sample. After the appropriate calibration with a set-up reference, such an empty cuvette, the absorption spectrum of the liquid sample can be calculated. At this point, the direct spectral algorithm computes the distance between a test spectrum and a known spectrum that represents each category or material. The test spectrum is assigned to the “closer” category, i.e. the one that offers the minimum distance. The known spectrum is usually represented by the average or mean spectrum of all the training spectra corresponding to the same category. Of course, this has its own statistical meaning, but we would like to propose a slightly more generalist approach.

If the spectrum acquisition is subject to measurement errors, the employment of the mean does not provide as much information as required. We propose to represent each category or material by a line with a specific shape and thickness. The line *shape* is truly related with the spectral variation of the absorbance, whereas the *thickness* is related with the statistics underlying the measurements, and not only the mean. The mean is usually employed because Gaussian probability density function (pdf) of the measurements is assumed. If we change the mean for the median and if we work with interquartile range (IRQ) or with percentile (PC) ranges to delimit the thickness of the representative mark of each category, no explicit assumption of the data pdf is made. This way of operation has also a beneficial consequence, the automatic and indirect removal of measurement outliers. Figure 1 represents the block diagram followed to obtain each spectral mark.

Once the spectral spots are obtained, they are transformed into images. By a simple AND operation between the images of each data category and the image corresponding to the test sample we can easily and time efficiently obtain the associated category of the unknown spectrum. This procedure is graphically displayed in Figure 2.

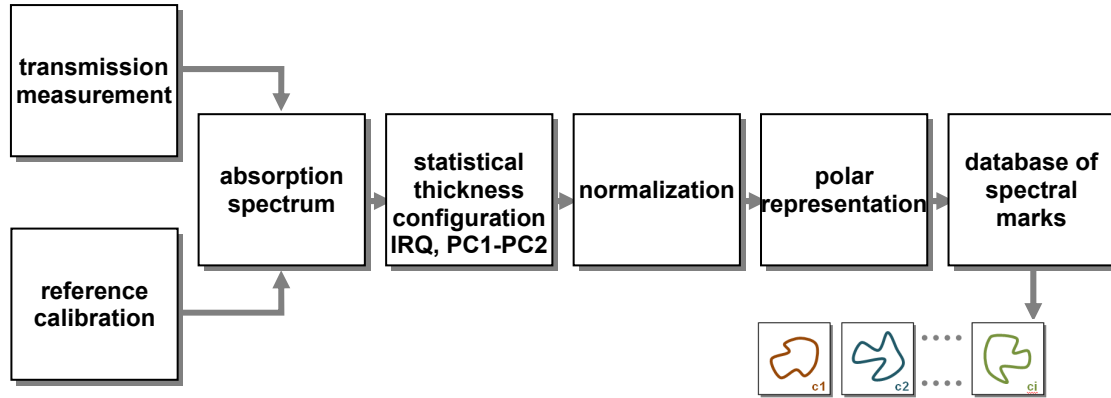


Fig. 1. Spectral data processing to obtain the “spectral marks” in the training stage.

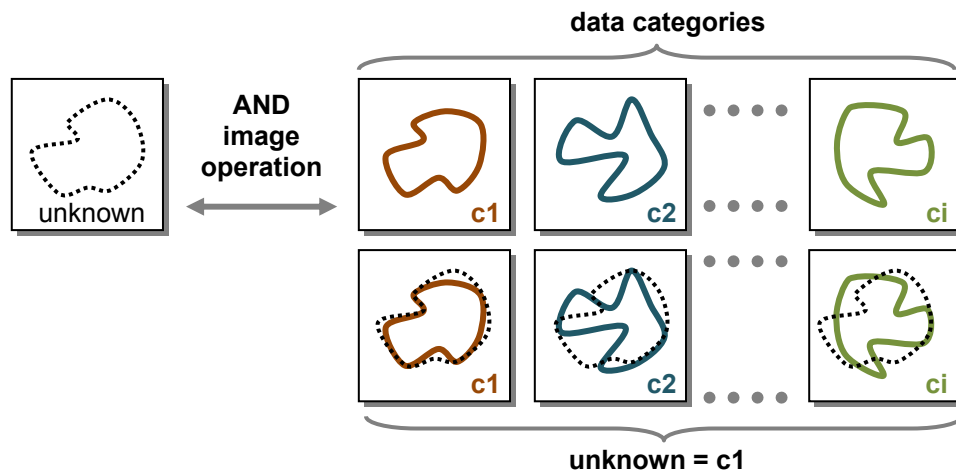


Fig. 2. “Spectral spot” testing stage.

3. EXPERIMENTAL SET-UP

Optical fiber absorption spectroscopy in the ultraviolet, visible and near infrared regions (specifically, from 200 nm to 1100 nm) was performed to characterize the spectra of liquid samples. For that purpose, the setup shown in Figure 3 was devised. As can be seen in the figure, light from a deuterium tungsten halogen source from Ocean Optics, model DH-2000, was coupled to a cuvette holder. As a proof of principle of the proposed discriminant analysis, different samples of oils from seeds (flax, grapeseed, sunflower, sesame) and olives (arbequina, picual, cornicabra, hojiblanca) were placed inside the cuvette holder in disposable plastic cuvettes of 0.5 cm lightpath. The transmitted light through the system was measured using a high-resolution spectrometer (Ocean Optics, HR2000+CG). Finally, solarization resistant optical fibers, with 600 μm of core diameter, were used to couple light between the source and the cuvette holder and the cuvette holder and the spectrometer. The system was calibrated using an empty cell, as blank reference, and a dark reference, which was taken when the light path to the spectrometer was blocked.

4. RESULTS AND DISCUSSION

Eight different types of oils have been evaluated to obtain their spectral plots. A hundred absorbance spectra in the 200-1100 nm range have been measured for each oil type in different days and conditions. Figure 4 shows the corresponding spectral marks for each oil type considering that the mark thickness is fixed by the IQR of the spectral measurements, i.e. the statistical dispersion of the first and third quartile data. As it can be deduced from Figure 4, even within olive oil,

slightly differences in the spectral marks can be observed. Figure 5 shows the superimposed spectral marks where the differences become more evident.

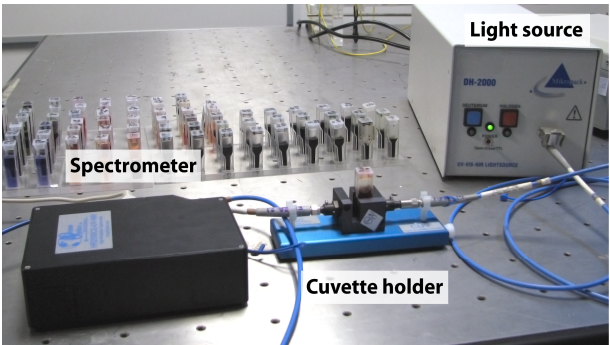


Fig. 3. Experimental setup used for liquids spectra characterization using optical fiber absorption spectroscopy.

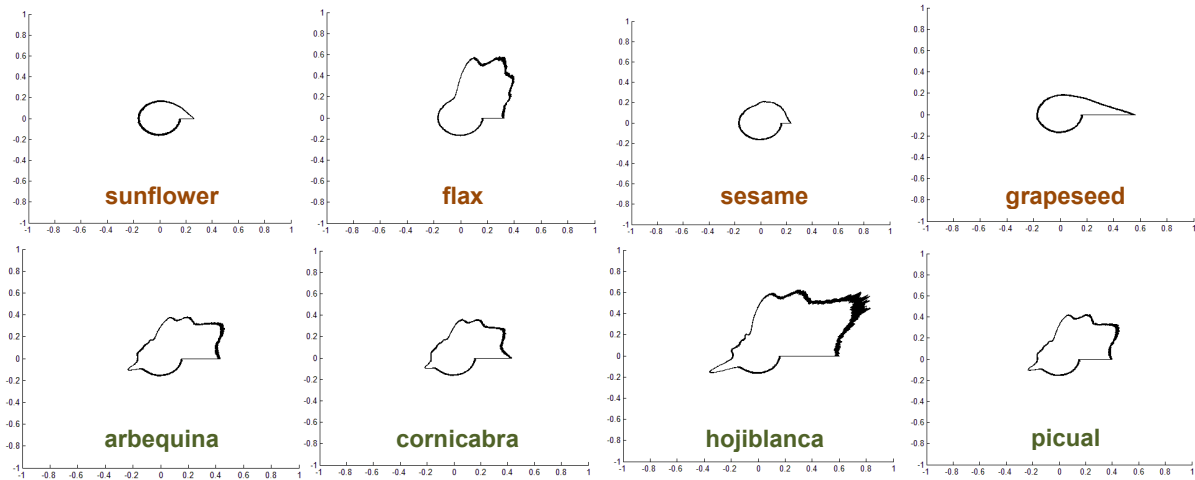


Fig. 4. “Spectral spots” of oils: seeds oils (brown-first row) and olive oils (green-second row).

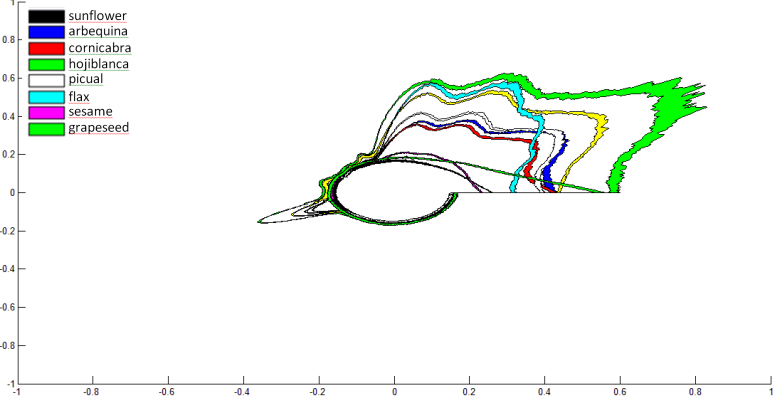


Fig. 5. Superimposed “spectral marks” of oils for a given IQR thickness.

In order to assess the algorithm performance, a three-step leave-one-out procedure has been implemented. The corresponding joined confusion matrix is generated and displayed in Table 1. This table represents the percentage of successful classification as the average of the results of the three steps of the cross-validation procedure. As can be seen, the algorithm behaves pretty well for the olive oils and show slightly worse results for the seeds oils.

Table 1. Confusion matrix of oils discrimination based on “spectral marks”.

	sunflower	flax	sesame	grapeseed	arbequina	cornicabra	hojiblanca	picual
sunflower	66,7	0	0	33,3	0	0	0	0
Flax	0	100	0	0	0	0	0	0
Sesame	16,7	0	83,3	0	0	0	0	0
grapeseed	16,7	0	0	83,3	0	0	0	0
arbequina	0	0	0	0	90,0	1,7	0	8,3
cornicabra	0	0	0	0	7,5	92,5	0	0
hojiblanca	0	0	0	0	0	0	100	0
picual	0	0	0	0	1,8	0	0	98,2

5. CONCLUSIONS

This work proposes and demonstrates the feasibility of a new algorithm for spectral qualitative discrimination analysis. The proposed methodology has been named “spectral marks” because it is based on the association of each material category to a specific figure, shape or mark directly related to the absorbance spectrum of the material. The width of each spectral spot is fixed taking into account the statistical spreading of the measurements and, therefore, no assumption on the data probability density function is made. Furthermore, an indirect treatment of the spectral outliers is also performed as data points outside the statistics defining the width of the spectral marks are removed. Apart from this, the main advantage of the proposal relies on its fast implementation. The classification of an unknown sample only needs to perform AND operations between a reduce set of images, each of them representing a material category. This procedure can be translated easily to hardware processing platforms (FPGA’s, DSP’s, etc.) and can also be parallelized to improve the sensor time response. The feasibility of the algorithm has been validated with the automatic discrimination of different oil samples (seed and olive) that have been measured with an optical fiber absorption spectroscopy set-up. Acceptable accuracy has been attained in the oils discrimination that can be improved after studying the best statistics to delineate the width of the spectral spot.

ACKNOWLEDGEMENTS

This work has been supported by the project TEC2007-67987-C02-01 and the measurement effort of Sergio Diego.

REFERENCES

1. Workman, J. Springsteen, A., [Applied Spectroscopy: a compact reference for practitioners], Academic Press, (1998).
2. Dale, J.M., Klatt, L., “Principal component analysis of diffuse Near-Infrared reflectance data from paper currency,” Applied Spectroscopy, 43, 1399-1405 (1989).
3. Joliffe, I.T., [Principal Component Analysis], Springer-Verlag, (1986).
4. Downey, G., Robert, P., Bertrand, D., Kelly, P.M., “Classification of commercial skim milk powders according to heat treatment using factorial discriminant analysis of Near-Infrared reflectance spectra,” Applied Spectroscopy, 44, 150-155 (1990).
5. Gemperline, P.J., Webber, L.D., Cox, F.O., “Raw-materials testing using soft independent modeling of class analogy analysis of near-infrared reflectance spectra,” Anal. Chem., 61, 138-144 (1989).
6. Drennen, J.K., Lodder, R.A., “Qualitative analysis by Near-Infrared spectrometry: a comparison of discriminant methods in dissolution testing,” Spectroscopy, 6, 34-39, (1991).
7. Burns, D. A. and Ciurczak, E. W., [Handbook of Near-Infrared Analysis], Third Edition. CRC press (2008).